

The background features abstract, overlapping green geometric shapes in various shades of lime and forest green, primarily located on the left and right sides of the slide. The central area is white, providing a clean space for the text.

# Capstone Project

The Battle of Neighborhoods

# Introduction

- ▶ Background : Safety is a top concern when moving to a new area. If you don't feel safe in your own home, you're not going to be able to enjoy living there.
- ▶ Problem : This project aims to select the safest borough in London based on the total crimes, explore the neighborhoods of that borough to find the 10 most common venues in each neighborhood and finally cluster the neighborhoods using k-mean clustering.
- ▶ Interest : Expats who are considering to relocate to London will be interested to identify the safest borough in London and explore its neighborhoods and common venues around each neighborhood.

# Data Acquisition and Cleaning

**Data Acquisition :** The data acquired for this project is a combination of data from three sources.

- ▶ The first data source of the project uses a London crime data that shows the crime per borough in London.
- ▶ The second source of data is scraped from a wikipedia page that contains the list of London boroughs .
- ▶ The third data source is the list of Neighborhoods in the Royal Borough of Kingston upon Thames as found on a wikipedia page.

**Data Cleaning :** The Data cleaning process for each of the data sources are done separately.

- ▶ From the London crime data, the crimes during the most recent year (2016) are only selected.
- ▶ The second data is scraped from a wikipedia page using the **Beautiful Soup** library in python. Using this library we can extract the data in the tabular format as shown in the website.
- ▶ The two datasets are merged on the Borough names to form a new dataset that combines the necessary information in one dataset
- ▶ After visualizing the crime in each borough we can find the borough with the lowest crime rate and hence tag that borough as the safest borough.
- ▶ The new dataset is used to generate the 10 most common venues for each neighborhood using the Foursquare API, finally using k means clustering algorithm to cluster similar neighborhoods together.

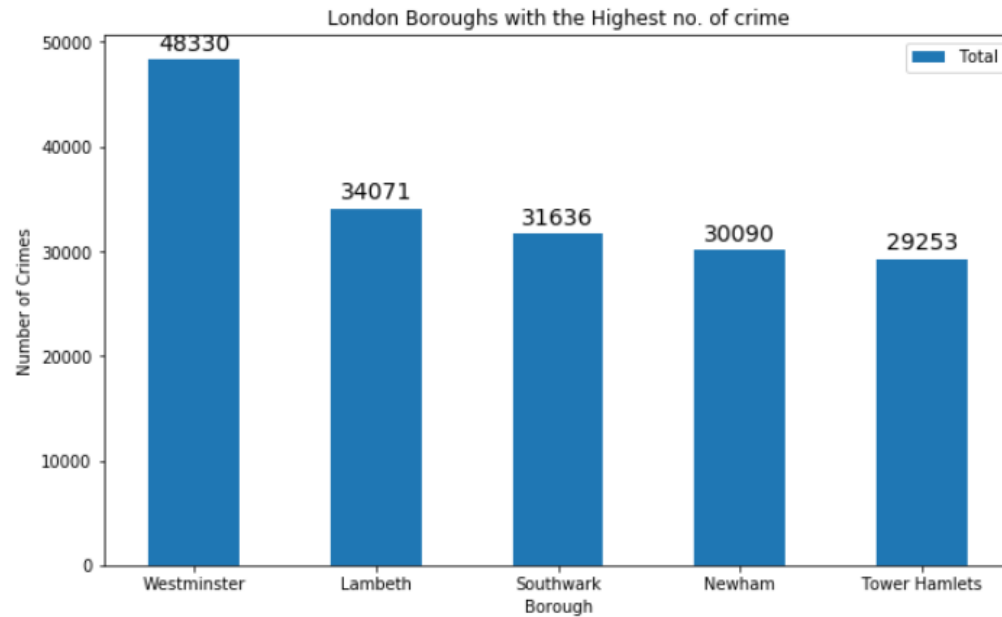
# Methodology

## ► Exploratory Data Analysis

	Burglary	Criminal Damage	Drugs	Other Notifiable Offences	Robbery	Theft and Handling	Violence Against the Person	Total
count	33.000000	33.000000	33.000000	33.000000	33.000000	33.000000	33.000000	33.000000
mean	2069.242424	1941.545455	1179.212121	479.060606	682.666667	8913.121212	7041.848485	22306.696970
std	737.448644	625.207070	586.406416	223.298698	441.425366	4620.565054	2513.601551	8828.228749
min	2.000000	2.000000	10.000000	6.000000	4.000000	129.000000	25.000000	178.000000
25%	1531.000000	1650.000000	743.000000	378.000000	377.000000	5919.000000	5936.000000	16903.000000
50%	2071.000000	1989.000000	1063.000000	490.000000	599.000000	8925.000000	7409.000000	22730.000000
75%	2631.000000	2351.000000	1617.000000	551.000000	936.000000	10789.000000	8832.000000	27174.000000
max	3402.000000	3219.000000	2738.000000	1305.000000	1822.000000	27520.000000	10834.000000	48330.000000

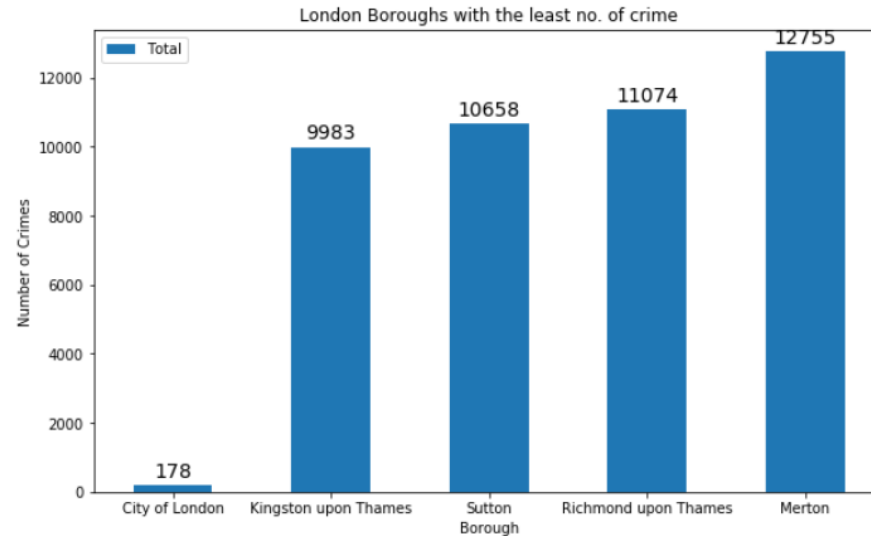
- The count for each of the major categories of crime returns the value 33 which is the number of London boroughs. 'Theft and Handling' is the highest reported crime during the year 2016 followed by 'Violence against the person', 'Criminal damage'. The lowest recorded crimes are 'Drugs', 'Robbery' and 'Other Notifiable offenses'.

# Boroughs with High Crime rates



- ▶ Comparing five boroughs with the highest crime rate during the year 2016 it is evident that Westminster has the highest crimes recorded followed by Lambeth, Southwark, Newham and Tower Hamlets. Westminster has a significantly higher crime rate than the other 4 boroughs

# Boroughs with low Crime rates



- ▶ Comparing five boroughs with the lowest crime rate during the year 2016, City of London has the lowest recorded crimes followed by Kingston upon Thames, Sutton, Richmond upon Thames and Merton.
- ▶ City of London has a significantly lower crime rate because it is the 33rd principal division of Greater London but it is not a London borough. It has an area of 1.12 square miles and a population of 7000 as of 2013 which suggests that it is a small area.
- ▶ we will consider the next borough with the lowest crime rate as the safest borough in London which is Kingston upon Thames.

# Modelling

- Using the final dataset containing the neighborhoods in Kingston upon Thames along with the latitude and longitude, we can find all the venues within a 500 meter radius of each neighborhood by connecting to the Foursquare API.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Berrylands	51.393781	-0.284802	Surbiton Racket & Fitness Club	51.392676	-0.290224	Gym / Fitness Center
1	Berrylands	51.393781	-0.284802	Alexandra Park	51.394230	-0.281206	Park
2	Berrylands	51.393781	-0.284802	K2 Bus Stop	51.392302	-0.281534	Bus Stop
3	Canbury	51.417499	-0.305553	Canbury Gardens	51.417409	-0.305300	Park
4	Canbury	51.417499	-0.305553	The Boater's Inn	51.418546	-0.305915	Pub

- One hot encoding is done on the venues data. The Venues data is then grouped by the Neighborhood and the mean of the venues are calculated, finally the 10 common venues are calculated for each of the neighborhoods.
- To help people find similar neighborhoods in the safest borough we will be clustering similar neighborhoods using K - means clustering which is a form of unsupervised machine learning algorithm that clusters data based on predefined cluster size. We will use a cluster size of 5 for this project that will cluster the 15 neighborhoods into 5 clusters.



# Results

## ► Cluster 1

	Neighborhood	Borough	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue
1	Canbury	Kingston upon Thames	51.417499	-0.305553	0	Pub	Tennis Court	Shop & Service	Spa	Fish & Chips Shop	Supermarket	Indian Restaurant	Plaza
4	Hook	Kingston upon Thames	51.367898	-0.307145	0	Indian Restaurant	Bakery	Supermarket	Fish & Chips Shop	Women's Store	Fast Food Restaurant	Department Store	Dessert Shop
5	Kingston upon Thames	Kingston upon Thames	51.409627	-0.306262	0	Coffee Shop	Italian Restaurant	Clothing Store	Café	Pub	Bakery	Department Store	Sandwich Place
9	New Malden	Kingston upon Thames	51.405335	-0.263407	0	Gym	Bar	Korean Restaurant	Chinese Restaurant	Supermarket	Sushi Restaurant	Office	Indian Restaurant
10	Norbiton	Kingston upon Thames	51.409999	-0.287396	0	Platform	Food	Pub	Italian Restaurant	Indian Restaurant	Bagel Shop	Coffee Shop	Fried Chicken Joint
12	Seething Wells	Kingston upon Thames	51.392642	-0.314366	0	Indian Restaurant	Café	Pub	Coffee Shop	Harbor / Marina	Gym	Gym / Fitness Center	Hotel
13	Surbiton	Kingston upon Thames	51.393756	-0.303310	0	Coffee Shop	Pub	Pharmacy	Grocery Store	Italian Restaurant	Gastropub	Bistro	Post Office

- The cluster one is the biggest cluster with 7 of the 15 neighborhoods in the borough Kingston upon Thames. Upon closely examining these neighborhoods we can see that the most common venues in these neighborhoods are Restaurants, Pubs, Cafe, Supermarkets, and stores.

## ► Cluster 2

	Neighborhood	Borough	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue
2	Chessington	Kingston upon Thames	51.358336	-0.298622	1	Construction & Landscaping	Women's Store	Fast Food Restaurant	Department Store	Dessert Shop	Discount Store	Donut Shop	Dry Cleaner	Electronics Store

- The second cluster has one neighborhood which consists of Venues such as Restaurants, Golf courses, and wine shops.

## ► Cluster 3

	Neighborhood	Borough	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue
0	Berrylands	Kingston upon Thames	51.393781	-0.284802	2	Gym / Fitness Center	Park	Bus Stop	Deli / Bodega	Department Store	Dessert Shop	Discount Store	Donut Shop	Dry Cleaner
8	Motspur Park	Kingston upon Thames	51.390985	-0.248898	2	Gym	Park	Restaurant	Bus Stop	Soccer Field	Women's Store	Falafel Restaurant	Dessert Shop	Discount Store

- The third cluster has two neighborhood which consists of Venues such as Train stations, Restaurants, and Furniture shops.

## ► Cluster 4

	Neighborhood	Borough	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue
7	Malden Rushett	Kingston upon Thames	51.341052	-0.319076	3	Grocery Store	Pub	Garden Center	Restaurant	Falafel Restaurant	Deli / Bodega	Department Store	Dessert Shop	Discount Store

- The fourth cluster has one neighborhoods in it, these neighborhoods have common venues such as Parks, Gym/Fitness centers, Bus Stops, Restaurants, Electronics Stores and Soccer fields etc.

## ► Cluster 5

	Neighborhood	Borough	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue
6	Kingston Vale	Kingston upon Thames	51.431850	-0.258138	4	Grocery Store	Bar	Sandwich Place	Soccer Field	Farmers Market	Department Store	Dessert Shop	Discount Store
11	Old Malden	Kingston upon Thames	51.382484	-0.259090	4	Grocery Store	Construction & Landscaping	Train Station	Food	Women's Store	Farmers Market	Department Store	Dessert Shop
14	Tolworth	Kingston upon Thames	51.378876	-0.282860	4	Grocery Store	Pharmacy	Coffee Shop	Bowling Alley	Restaurant	Sandwich Place	Bus Stop	Soccer Field

- The fifth cluster has three neighborhood which consists of Venues such as Grocery shops, Bars, Restaurants, Furniture shops, and Department stores.

# Discussion

- ▶ The aim of this project is to help people who want to relocate to the safest borough in London, expats can choose the neighborhoods to which they want to relocate based on the most common venues in it.
- ▶ If a person is looking for a neighborhood with good connectivity we can see that Clusters 3 and 4 have Train stations and Bus stops as the most common venues.
- ▶ If a person is looking for a neighborhood with stores and restaurants then the neighborhoods in the first cluster are suitable.
- ▶ For a family the neighborhoods in Cluster 5 are more suitable due to the common venues such as Parks, Gym/Fitness centers, Bus Stops, Restaurants, Electronics Stores and Soccer fields.

# Conclusion

- ▶ This project helps a person get a better understanding of the neighborhoods with respect to the most common venues in that neighborhood.
- ▶ It is always helpful to make use of technology to stay one step ahead i.e. finding out more about places before moving into a neighborhood.
- ▶ We have just taken safety as a primary concern to shortlist the borough of London. The future of this project includes taking other factors such as cost of living in the areas into consideration to shortlist the borough based on safety and a predefined budget.